DOCUMENT RESUME

ED 321 525 FL 018 563

AUTHOR Rubin, Donald L.
TITLE Sociolinguistic Test Item Review.
SPONS AGENCY Auburn Univ., Montgomery, AL. Center for Business and Economic Development.
PUB DATE Jun 89
NOTE 6p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Black Dialects; Dialect Studies; English; *Item Bias; Language Research; *Multiple Choice Tests; *Nonstandard Dialects; Testing; *Test Items; *Test Validity

ABSTRACT

Because the language of a multiple choice test is formal and often unfamiliar, certain linguistic features may lead a test-taker to misconstrue the test instructions, questions, or answers. When this happens, a shared understanding of meaning between tester and test-taker is not present, and the test results are invalid. Although this problem exists for all test-takers, it is more acute for members of non-mainstream speech communities because they are less likely to suspend their normative expectations of discourse. This study was undertaken to identify potential sources of linguistic distortion that might undermine a multiple-choice test's validity for measuring job-relevant knowledge, skills, and abilities among Black English Vernacular speakers in a southern urban setting. Problematic linguistic features identified include: undue syntactic complexity, tense switching, the use of incomplete question stems, plural/possessive ambiguity, and others. Included is a bibliography of significant references on multiple-choice test bias, and a form that can be used when examining test items for sociolinguistic bias. (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED321525

SOCIOLINGUISTIC TEST ITEM REVIEW

Donald L. Rubin, Ph.D.
Departments of Language Education
and Speech Communication
The University of Georgia
Athens, GA 30602

on behalf of
Center for Business and Economic Development
Auburn University at Montgomery

Project Summary

A variety of factors can introduce discrepancies between test-takers' observed scores on a test and their ideal "true scores." Taking a test is a type of communication event. Most often, written language is the medium of that communication. The phrasing of the test instructions and the test items are intended to tap the test-takers trait-relevant or job-relevant knowledge. The language of the test-takers' responses, often supplied in multiple choice foils, ought to match the test-taker's intended meaning. To the extent that communication fails--to the extent that questions are likely to be misconstrued and answers are likely to misrepresent--distortion is introduced that is extraneous to the trait to be measured.

sociolinguistic test reviewing. Some such linguistic distortion is inevitable for nearly all test-takers. After all, the language of formal testing is hardly the most familiar language for any speaker. On the other hand, that distortion is not uniformly distributed among all social groups. The purpose of a sociolinguistic test review is to mitigate distortion in scores that results because certain groups of test-takers belong to non-mainstream language communities. Non-mainstream language communities equip their members with divergent rules for construing the way tests work, the meaning of test questions, and the meaning of supplied test responses. Validity is not a property of tests, per se, but rather an attribute of the way tests are used for specific populations of test-takers. In this sense, sociolinguistic test reviewing is a crucial component of the validation process.

For the project at hand, the purpose of the sociolinguistic test review was to identify potential sources of linguistic distortion that might undermine the tests' validity when used to measure job-relevant knowledge, skills, and abilities among core speakers of Black English Vernacular (BEV) in a Southern urban setting.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Rubin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL018563

BEST COPY AVAILABLE

establishing sociolinguistic criteria for test reviewing. Reviewing test content for general cultural bias is a well established practice. It is relatively rare, however, for test constructors to conduct item reviews that focus on sociolinguistic factors in particular. A search of research and development literature pertaining to the topics of cultural bias in testing, dialect interference in comprehension, cultural factors in test-taking, and related topics yielded approximately 130 abstracts of interest. Approximately 25 documents were ultimately reviewed. The most relevant of these are cited in Appendix A to this summary.

Issues of sociolinguistic concern include (1) direct dialect interference, (2) unfamiliar communication events, and (3) unfamiliar discourse norms.

With respect to dialect interference, 25 years of research confirm that **reading comprehension** per se is not related to BEV speech. On the other hand, some specific information processing strategies may be cued in ways that are peculiar to BEV. These are manifest in different meanings associated with relationship terms like prepositions and comparatives. Also, particular lexical items may have nonstandard definitions. A further issue related to the impact of direct dialect interference pertains to questions which test job-related uses of Standard Edited English forms. The research on dialect influences in writing indicates that BEV speakers are unlikely to reproduce in writing the syntax of their spoken dialect (e.g., multiple negation). But dialect effects on writing are common at the level of morphology: pronoun and verb inflections, tense and possessive markings. If test constructors are committed to including items that demand recognition of Standard Edited English morphology, they must carefully evaluate what weight these items ought to be accorded.

Tests constitute a special kind of communication event. Unlike conversational questions, test questions are not requests for information unknown to the asker. Instead, they are demands for test-takers to display knowledge already known to the asker. Members of mainstream speech communities are more familiar with related quasi-questioning routines. For members of non-mainstream communities, and for BEV speakers in particular, however, the language behaviors of test-taking are relatively alien. Related sociolinguistic criteria for cultural fairness, therefore, suggest that tests be untimed **power tests**. Similarly, test questions must explicitly state the kinds of responses demanded of test-takers. Factors that unnecessarily complicate the testing event, such as failing to repeat information that may be crucial to a series of related questions, ought to be minimized. For the same reason, questions that are phrased in unduly complex syntax may contribute to sociolinguistic bias.

Because standardized testing is an unusual communication event, it incorporates discourse norms that are differentially familiar to members of different speech communities. Examples of such discourse norms include negative questions ("which of the following is not..."), distractors like "all of the above" and "none of the above," and incomplete questions stems. By the same token, test questions sometimes include excess or irrelevant information that may not be needed to solve the problem posed. Part of what it means to be "test wise" is to handle such violations of **maxims of quantity and quality**. But members of non-mainstream speech communities may be less likely to approach a testing situation by "appropriately" suspending their normative expectations of discourse.

The sociolinguistic criteria for reviewing test items are summarized on the evaluation form, which appears in Appendix B.

conducting the sociolinguistic item review. None of the sociolinguistic criteria identified is proven to specifically undermine BEV speakers' test performance. Most likely, no single feature exerts a measurable independent effect. Instead, sociolinguistic features tend to act in concert, as clusters of co-occurring language variables each contributing to communication outcomes. Nevertheless, the most conservative course of action is to flag each occurrence of an identified feature as a defect. Some tolerance for minor defects can be accepted, however. For example, if an item contains unduly complex syntax, it makes sense to revise it from the perspective of readability. But from the perspective of sociolinguistics, complex syntax alone may not be a fatal flaw. On the other, violations of discourse norms--defects like negatively phrased questions or incomplete question stems--do demand that an item be revised or rejected.

APPENDIX A

Key References

Boldt, R.F., Levin, M.K., Powers, D.E., Griffin, M., Troike, R. C., Wolfram, W., & Ratliff, F.R. (1977). Sociolinguistic and measurement considerations of armed services selection batteries. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Heath, S.B. (1982). Questioning at home and at school: A comparative study. In G. Spindler (Ed.), Doing ethnography: Educational ethnography in action (pp. 105-131). New York:_

Labov, W. (1976). Systematically misleading data from test questions. Urban Review, 9, 146-169.

McPhail, I.P. (1975). Overcoming dialect problems on the S.A.T.: A descriptive study of a program for urban black high school students. Philadelphia: University of Pennsylvania Reading Clinic. [EDRS No. ED103872].

Orr, E.W. (1987). Twice as less: Black English and the performance of black students in mathematics and science. New York: W.W. Norton.

Williams, R.L. (1975). Developing cultural specific assessment devices: An empirical rationale. In R.L. Williams (Ed.), Ebonics: The true language of black folks (pp. 110-132). St. Louis: Robert L. Williams and Associates.

Williams, R.L., & Rivers, L.W. (1975). The effects of language on the test performance of black children. In R.L. Williams (Ed.), Ebonics: The true language of black folks (pp. 96-109). St. Louis: Robert L. Williams and Associates.

Williams, T.E., & Boyd, B.H. (March, 1982). Attractiveness of "Black English" foils: An examination of a potential source of item bias. Paper presented at the Annual Meeting of the American Educational Research Association, New York. [EDRS No. ED222581].

Wolfram, W. (1976). Levels of sociolinguistic bias in testing. In D.S. Harrison and T. Trabasso (Eds.), Black English: A seminar (pp. 265-287). Hillsdale, NJ: Lawrence Erlbaum.

APPENDIX B:

SOCIOLINGUISTIC ITEM REVIEW

Item No: _____

DIAGNOSES

undue syntactic complexity: ____clausal embedding
____reduced clauses

incomplete question stem_____

negative item_____

tense switching_____

subjunctive mood_____

modal verbs ____CAN=>COULD ____ WILL=>WOULD ____others:

plural/possessive ambiguity ____

relative pronoun ambiguity: ___conjunctive WHICH
____ subordinating OF ___human WHICH

problematic comparatives: ____ X-AS-Y-AS-Z ____X-ER-OF-Y
____THAN => AS

negative + quantifier ____NOT ANY ___NOT ALL ___NOT EACH
____NOT EVERY ____ others:

problematic prepositions ___ behind => IN BACK OF ___ AT => TO
____ TO => AT, IN, ON, ONTO ____ others:

EXISTENTIAL THERE=>IT____

lexical items: ___WHOLE ____STARTING ____HALF ____others:

low frequency or unpredictable words as homophones of higher probability words ____

remarks:

RECOMMENDATIONS

___accept item ___discard item

___revise stem ____revise foils ___a ___b ___c ___d